

A STUDY OF QUASI-EXPERIMENTAL CONTROL GROUP METHODS FOR ESTIMATING POLICY IMPACTS

by

W. Robert Reed and Cynthia Rogers*
Department of Economics
University of Oklahoma
Norman, OK 73019

Abstract

This study examines the efficacy of Quasi-Experimental Control Group (QECG) methods for estimating policy impacts. It establishes that QECG estimators can outperform the conventional regression (CR) estimator when policy adoption is endogenous (nonrandom), the relationship between outcomes and policies is nonlinear, and CR equations do not correctly specify the nonlinear form of the relationship. In the case of perfect matching, QECG methods produce unbiased estimates. In the case of imperfect matching, QECG estimators will be biased. To address this and other issues, we develop a more general QECG estimator that (1) allows control places to be matched to more than one treatment place, and (2) weights observations by the “closeness” of the match. Using Monte Carlo analysis we demonstrate that our estimator substantially improves estimates of policy impacts.

JEL Classification: R58, C20, C15

May 4, 2001

*The contact author is Cynthia Rogers, email: crogers1@ou.edu, fax: (405) 325-5842, phone: (405) 325-5843.

Acknowledgements: Dan Sutter made a number of valuable comments that improved our understanding of the differences between quasi-experimental and cross sectional-time series regression models. Financial assistance from the Oklahoma Office of State Finance is gratefully acknowledged. The opinions and conclusions stated in this paper are exclusively those of the authors and are not necessarily endorsed by the Oklahoma Office of State Finance.

1. INTRODUCTION

Over the past decade there has been increasing interest in using quasi-experimental control group (QECG) methods to estimate the impact of state and local development policies. This interest stems from a desire to circumvent a number of challenging problems associated with conventional regression (CR) analyses. Prominent among these problems are (i) the endogeneity of policy adoption and (ii) misspecification (Bartik, 1991, pages 30-36; Bartik, 1997, pages 68-69; and Wasylenko, 1997, page 3).

The appeal of QECG analysis is its attempt to replicate traditional experimental design in the evaluation of public policy.¹ QECG methods have four components. First, a "treatment" is chosen for study. The treatment is usually a state or local development policy, such as a public infrastructure investment (e.g., a highway, dam, or bypass) or tax policy (e.g., an employment subsidy or enterprise zone). Second, one or more "treatment places" are identified. These are geographical places where the impact of the public policy is located (e.g., the location of new highway construction, or the locality in which the specified tax treatment is being applied). Third, one or more "control places" are selected. These are geographical places that closely match a given treatment place on a number of pre-selected characteristics. The fourth component consists of a post-treatment comparison of outcome variables—usually related to population, employment

¹ Isserman and Merrified (1987, page 3) write,

"The traditional experimental research design...generally is not possible in regional and other place-related research. The quasi-experimental design is feasible, however. A group of places where something did not occur (no treatment) can be selected nonrandomly to be a control group for a place or places where something did occur (treatment). In a sense, aspects of a laboratory experiment are reconstructed after the treatment has occurred--hence, the term quasi-experimental."

or income—between the treatment and control group places.² Quantifiable differences in the outcome variables are attributed to the treatment and interpreted as the “impact” of the given policy.

Central to QECG analysis is the matching of treatment with control places. While there are many variants, most QECG studies employ either a case study or a “twins” approach.³ Case studies generally match a treatment place to multiple control places. Examples are Greenburg et al. (1998), Isserman and Merrifield (1982), and Isserman and Merrifield (1987)). In case studies, the outcome variable for the treatment place is compared to the mean (or, occasionally, the median) of the outcome variable for the set of control places. While the case study approach is amenable to analyzing more than one treatment place, these analyses are usually conducted separately from each other. That is, multiple analyses are generally not combined to produce a summary statistic of impact. Further, the analyses are usually qualitative in the sense that no formal hypothesis tests are carried out.⁴

² Some studies employ a pre-test that compares outcome variables between the treatment and controls in a period prior to the post-test years but after the matching period. In this case, however, the similarity in the pre-test period does not necessarily increase confidence about continued similarity in the post-treatment period between the treatment and control observations. See Wojan and Bailey (1998) for a discussion.

³ Bohm and Lind (1993) use differences in the means of outcome variables for treatment and control places, among other approaches, in their analysis of regional employment subsidies in Sweden. In a related approach using time series data, Broder et al. (1992) regress the outcome variable for the treatment place on the outcome variable for the matched control place(s), a time trend, and a dummy variable indicating the time period in which the treatment was applied.

⁴ Isserman and Beaumont (1989) propose point estimates and corresponding statistical tests for evaluating treatment impacts.

In contrast, "twins" studies examine a plurality of treatment places, matching each treatment place to a single control place. Examples of twins studies are Rephann and Isserman (1994), Bohm and Lind (1994), Isserman and Rephann (1995), Rephann et al. (1997), and Aleseyed et al. (1998). Treatment impact is estimated as the mean of the pairwise differences in the respective outcome variable. The statistical framework of paired sampling is employed to formally test the null hypothesis of no impact. An important assumption implicit in the paired sample framework is that observations are independently and identically distributed.

One implication of the "independently distributed" assumption is that observations cannot share the same control places. In other words, control places are required to be unique to a particular treatment place. The requirement of uniqueness is costly. If a control place is the "best" match for more than one treatment place, some procedure must be used to assign the control county to one, and only one, treatment place.⁵ This means that only one of the treatment places will be paired with its best match. The other treatment places will be paired with control places that are their second, third, or lower "best" match. Thus, the independence assumption, with its requirement of uniqueness, interferes with the goal of pairing each treatment county with its best match.

For researchers interested in estimating state and local policy impacts, the advantages of QECG methods over CR methods have not been firmly established. While

⁵ Methods for selecting unique control places for each treatment place range from informal matching to formal algorithms. An example of the latter is Rosenbaum (1989).

endogeneity and misspecification present difficult problems for CR methods, it is not clear why one would expect QECG methods to produce better estimates.

Furthermore, as suggested by the preceding discussion, there are many questions about the best way to implement QECG methods.

1. Should control places be matched to only one treatment place (thus imposing the requirement of uniqueness)? Or should control places be permitted to be matched to more than one treatment place?
2. Should observations be weighted equally, as is typically done in QECG analysis? Or should observations be weighted by the “closeness” of the match?
3. Should treatment places be matched to only one control place (i.e., pairwise matching)? Or is it better to match each treatment place to multiple control places?

Finally, QECG methods borrow their conceptual framework from traditional experimental design where treatment and control observations can be very closely--if not perfectly--matched. In contrast, when it comes to place-related policy impacts, QECG methods must resort to “imperfect matching”. The consequences of imperfect matching for the properties of QECG methods have not been carefully studied.

This study addresses these issues through a combination of analytical and experimental studies. Section 2 identifies how endogeneity and misspecification combine to cause CR estimates to be biased. It goes on to demonstrate how QECG analysis can overcome these problems to produce unbiased estimates when treatment and control places are “perfectly matched.” Section 3 generalizes the analysis of QECG methods to the more realistic case when matching of treatment and control places is “imperfect.” We show that QECG estimators will no longer be unbiased when the assumption of perfect matching is invalid. However, they may still perform better than the CR estimator. We develop a new QECG estimator that (i) allows control places to be

matched to more than one treatment place (thus ensuring that the best match is always used), and (ii) weights observations on the basis of the closeness of matches. Section 4 uses Monte Carlo analysis to compare the absolute and relative performances of QECG and CR estimators in a variety of Monte Carlo experiments. The experiments demonstrate that QECG estimators in general, and our new QECG estimator in particular, can substantially improve estimation of policy impacts. Section 5 concludes.

2. QECG ESTIMATION WHEN MATCHING IS PERFECT

This section illustrates the potential advantages of QECG methods. When the adoption of policies by places is nonrandom and the empirical model is misspecified, CR estimates will be inconsistent. In contrast, QECG methods can produce unbiased estimates even if the functional form of the true model is unknown, as long as the treatment places are matched with identical control places.

Suppose policy-makers are interested in the impact of a particular policy on a population, employment, or income variable, y . There exist N observations of y , as well as D , a dummy variable indicating whether a given observation was “treated” with the policy, and \mathbf{X} , a set of K variables known to be independently related to y . The true model relating y to D and \mathbf{X} is given by

$$(1) \quad y_i = \alpha + \beta D_i + f(\mathbf{X}_i) + \varepsilon_i, i=1,2,\dots,N,$$

where ε is an i.i.d. error term, $E(\varepsilon)=0$, $E(\mathbf{D}'\varepsilon)=0$, $E(f(\mathbf{X})'\varepsilon)=0$, and f is a nonlinear function. In addition, $Cov(D, f(\mathbf{X})) \neq 0$; that is, the places receiving treatment are a nonrandom sample of places, or stated differently, the treatment variable is endogenous.

If a researcher could estimate equation (1) directly, conventional regression (CR) methods would produce consistent estimates of β .⁶ That is, even though the adoption of the policy is endogenous, as long as all the relevant variables are included in equation (2), the associated endogeneity would not be expected to produce bias (though it would generate multicollinearity).

Let us assume that the functional form of $f(\mathbf{X})$ is unknown. The researcher attempts to estimate the true model using the linear specification

$$(2) \quad y_i = \alpha + \beta D_i + \mathbf{X}_i \boldsymbol{\Gamma} + \nu_i, \quad i=1,2,\dots,N,$$

where $\boldsymbol{\Gamma}$ is a vector of coefficients. This introduces misspecification error if the true specification is nonlinear.⁷ Comparison of (1) with (2) shows that

$$(3) \quad \nu_i = \varepsilon_i + f(\mathbf{X}_i) - \mathbf{X}_i \boldsymbol{\Gamma}.$$

Let b be the OLS estimator of β in equation (2), and \mathbf{D}_* the vector of residuals from regressing D on the other independent variables in equation (2). If $\text{plim} \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) \neq 0$,

then $\text{plim} \left(\frac{\mathbf{D}_*' \boldsymbol{\nu}}{N} \right) \neq 0$ and $\text{plim}(b) \neq \beta$. In other words, if D is correlated with $f(\mathbf{X})$,

⁶ This assumes, of course, that $D_i \neq f(\mathbf{X}_i)$, in which case equation (2) would be characterized by perfect multicollinearity.

⁷ We restrict our analysis to the case where misspecification consists of functional form misspecification ($f(\mathbf{X}_i) \neq \mathbf{X}_i \boldsymbol{\Gamma}$). That is, we assume the correct set of variables \mathbf{X} are included in equation (2), but that the functional relationship between \mathbf{X} and y has been incorrectly specified. Misspecification can also occur through the inclusion of irrelevant variables or the omission of relevant variables. We empirically study the inclusion of irrelevant variables below. The omission of relevant variables is beyond the scope of this study.

and the functional form of f is misspecified in the regression equation, then D_* will be correlated with the regression equation's error term. As a result, CR analysis will produce inconsistent estimates of the policy impact, β .

We now show how QECG methods can be used to produce unbiased estimates of β in this case. Suppose each place receiving the treatment is matched to P places which (i) did not receive the treatment and (ii) are characterized by the exact same values of X . Let T_i identify the i th treatment place, and $C_{i,p}$, $p=1,2,\dots,P$, the P observations that exactly match this treatment observation (i.e., the i th treatment place's control group).

From equation (1) it follows that

$$(4) \quad y_{T_i} - \frac{\sum_{p=1}^P y_{C_{i,p}}}{P} = \beta + \left(f(X_{T_i}) - \frac{\sum_{p=1}^P f(X_{C_{i,p}})}{P} \right) + \left(\varepsilon_{T_i} - \frac{\sum_{p=1}^P \varepsilon_{C_{i,p}}}{P} \right).$$

Since $X_{T_i} = X_{C_{i,p}}$, $p=1,\dots,P$, the middle term on the right hand side of (4) drops out and

$$(5) \quad y_{T_i} - \frac{\sum_{p=1}^P y_{C_{i,p}}}{P} = \beta + \left(\varepsilon_{T_i} - \frac{\sum_{p=1}^P \varepsilon_{C_{i,p}}}{P} \right).$$

Define $Y_i = y_{T_i} - \frac{\sum_{p=1}^P y_{C_{i,p}}}{P}$ and suppose that T places are observed to have received the

treatment. Then least squares estimation of the regression model

$$(6) \quad Y_i = \beta + \text{error term}, \quad i=1,2,\dots,T,$$

yields an unbiased estimate of the policy impact, β .⁸ Note that this specializes to the twins approach commonly employed in QECG studies when $P=1$ and each control place is matched to only one treatment place.

Thus, QECG methods will be superior to CR methods when (i) the treatment variable is endogenous but all relevant variables are included in the analysis (so that the treatment variable is uncorrelated with the error term in the true model), (ii) the functional relationship between y and the explanatory variables is nonlinear and unknown, and (iii) each treatment place can be matched to a control place having identical characteristics (i.e., $X_{Ti} = X_{Ci,p}$, $p = 1, \dots, P$, for each of the T treatment counties). In practice, this latter condition is unlikely to be satisfied. The next section discusses the implications for the QECG estimator of β when this assumption is violated.

3. QECG ESTIMATION WHEN MATCHING IS IMPERFECT

The expected value of the QECG estimator given imperfect matching. As is well known, the search for a control group is unlikely to produce exact matches. If X contains variables that are continuous, exact matching will be improbable. Likewise, when the number of variables in X is sufficiently large relative to the number of places in the potential control group, it will be highly unlikely that any control place will match every characteristic of a given treatment place.

⁸ However, as we point out below, the estimation of the standard errors will be biased if control places are matched to more than one treatment place.

Given imperfect matching, one would like the control group to consist of places that most closely resemble the place receiving treatment. This requires a measure of “closeness.” One commonly used measure is the Mahalanobis distance measure (Rephann and Isserman, 1994; Isserman and Rephann, 1995), defined by

$$(7) \quad M_{i,j} = (\mathbf{X}_{Ti} - \mathbf{X}_j)' \mathbf{R}^{-1} (\mathbf{X}_{Ti} - \mathbf{X}_j);$$

where \mathbf{X}_{Ti} and \mathbf{X}_j are the vectors of characteristics associated with the i th treatment and j th non-treatment place, and \mathbf{R} is the variance-covariance matrix associated with the variables in \mathbf{X} .⁹ Multiplication by the inverse of the variance-covariance matrix has several attractive features. It serves to rescale differences in any given variable in terms of standard deviation units. It also prevents "double-counting" of differences in the case of positively correlated variables, and "canceling out" of differences when variables are negatively correlated.¹⁰

The Mahalanobis distance measure is useful for assigning a numerical value that measures how similar a given non-treatment place is to the i th treatment place. The smaller the value, the closer the match between the characteristics of the two places. The P non-treatment places with the smallest Mahalanobis values—i.e., the P places determined to be the closest matches--can then be selected as the control group for the i th treatment observation.

⁹ Mueser, Troske, and Gorislawsky (1999) use the Mahalanobis distance measure to match participants with nonparticipants in their analysis of Missouri’s Job-Training Partnership Act.

¹⁰ Equation (7) may look familiar as the χ^2 statistic corresponding to the hypothesis, $H_0 : \mathbf{X}_{Ti} = \mathbf{X}_j$, if observed place characteristics were assumed to be stochastic.

The true model relating y to D and \mathbf{X} was identified in equation (1) to be $y_i = \alpha + \beta D_i + f(\mathbf{X}_i) + \varepsilon_i$, $i=1,2,\dots,N$, where $E(\varepsilon)=0$, $E(\mathbf{D}'\varepsilon)=0$, $E(f(\mathbf{X})'\varepsilon)=0$, f is a nonlinear function, and $Cov(D, f(\mathbf{X})) \neq 0$. From there we derived the relationship in equation (4), which can be rewritten

$$(8) \quad y_{Ti} - \frac{\sum_{p=1}^P y_{Ci,p}}{P} = \beta + \omega_i, i = 1, 2, \dots, T;$$

$$\text{where } \omega_i = \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \right) + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right).$$

This can be expressed in matrix notation as

$$(9) \quad \mathbf{y} = \mathbf{i}\beta + \boldsymbol{\omega},$$

where \mathbf{y} , \mathbf{i} , and $\boldsymbol{\omega}$ are column vectors consisting of the T observations of

$$y_{Ti} - \frac{\sum_{p=1}^P y_{Ci,p}}{P}, \text{ a } T \times 1 \text{ vector of ones, and the } T \text{ realizations of the error term } \omega_i,$$

respectively.

The corresponding least squares estimator of β , call it $\hat{\beta}$, is given by

$$(10) \quad \hat{\beta} = (\mathbf{i}'\mathbf{i})^{-1} \mathbf{i}'\mathbf{y} = (\mathbf{i}'\mathbf{i})^{-1} \mathbf{i}'(\mathbf{i}\beta + \boldsymbol{\omega}) = \beta + \frac{\mathbf{i}'\boldsymbol{\omega}}{T}.$$

The expected value of this estimator is

$$(11) \quad E(\hat{\beta}) = \beta + E \left(\omega \left| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right. \right)$$

$$= \beta + E \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \middle| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right).^{11}$$

An important result for QECG analysis is that

$$E \left(\omega \middle| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right) \neq 0 \quad \text{when there is imperfect matching.} \quad \text{Since}$$

$\text{Cov}(D, f(\mathbf{X})) \neq 0$, it follows that

$$E \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \middle| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right) \neq 0. \quad \text{Thus}$$

$$(12.a) \quad E(\hat{\beta}) \neq \beta \quad \text{and}$$

$$(12.b) \quad \text{Bias}(\hat{\beta}) = E \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \middle| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right).$$

In words, when the assumption of perfect matching is violated, QECG methods will produce biased estimates of the policy impact. The nature of the bias is akin to sample selection bias and is easily understood through an example. Suppose D and $f(\mathbf{X})$ are positively correlated, so that a larger value of $f(\mathbf{X})$ makes it more likely that a given place will adopt the respective policy. Then

¹¹ Note that there are two nonrandom processes conditioning this product: (i) the nonrandom process that determines which places are treatment places, and (ii) the nonrandom process that determines which control places get matched to each treatment place.

$$E \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \middle| \begin{array}{l} \text{The nonrandom process} \\ \text{determining treatment} \\ \text{and control places} \end{array} \right) \text{ will also be greater than } 0, \text{ since}$$

treatment places will be characterized by larger values of $f(\mathbf{X})$, on average, than the control places they are matched with. In this case, QECG methods will tend to overestimate the size of the policy impact.

Crucial to this biasedness result is the assumption that $Cov(D, f(\mathbf{X})) \neq 0$. In contrast, if D were uncorrelated with $f(\mathbf{X})$, as would be the case if the process determining treatment and control places were random, the QECG estimator of β would be unbiased. But then CR estimates of β would also be unbiased, so that the *raison d'être* for using QECG methods would be voided.

A comparison of QECG and CR estimators. While QECG methods under imperfect matching will not produce unbiased—or for that matter, consistent—estimates of the treatment effect, they may still represent a substantial improvement over CR estimates. Let us compare the size of the biases produced by OLS estimation of equation (2) (representing CR analysis) and equation (8) (representing QECG analyses).

It is straightforward to show that the expected value of the OLS estimator of β in equation (2) is given by

$$(13.a) \quad E(b) = \beta + E \left(\left(\mathbf{D}_*' \mathbf{D}_* \right)^{-1} \mathbf{D}_*' f(\mathbf{X}) \right).$$

We can get to a more easily interpretable expression if we take the probability limit of b rather than its expectation. In this case,

$$(13.b) \quad plim(b) = \beta + \left(plim \left(\frac{\mathbf{D}_* \mathbf{D}_*'}{N} \right) \right)^{-1} plim \left(\frac{\mathbf{D}_* f(\mathbf{X})}{N} \right),$$

where $\mathbf{D}_* = \mathbf{M}_* \mathbf{D}$, $\mathbf{M}_* = \mathbf{I}_N - \mathbf{X}_* (\mathbf{X}_* \mathbf{X}_*)^{-1} \mathbf{X}_*'$, $\mathbf{X}_* = [\mathbf{i}_N \quad \mathbf{X}]$, and \mathbf{i}_N is a $N \times 1$ vector of ones. The associated asymptotic bias is given by

$$(13.c) \quad Bias(b) = \left(plim \left(\frac{\mathbf{D}_* \mathbf{D}_*'}{N} \right) \right)^{-1} plim \left(\frac{\mathbf{D}_* f(\mathbf{X})}{N} \right).$$

It is useful to note for illustrative purposes that when $Cov(\mathbf{D}, \mathbf{X}) = \mathbf{0}_{I \times K}$ (i.e., \mathbf{D} is related nonlinearly, but not linearly, to \mathbf{X}),

$$(14) \quad Bias(b) = \frac{Cov(\mathbf{D}, f(\mathbf{X}) - \mathbf{X}\boldsymbol{\Gamma})}{Var(\mathbf{D})}.^{12}$$

A comparison of equation (14) with equation (12.b) suggests that QECG methods are likely to be characterized by smaller biases than CR methods when

$$E \left(\left| \frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \right| \bullet \right) \text{ is small relative to } \left(plim \left(\frac{\mathbf{D}_* \mathbf{D}_*'}{N} \right) \right)^{-1} plim \left(\frac{\mathbf{D}_* f(\mathbf{X})}{N} \right).^{13}$$

Accordingly, factors that are likely to affect the absolute and relative performances of the CR and QECG estimators are (i) the closeness of matches between treatment and control

¹² Of course, when $Cov(\mathbf{D}, \mathbf{X}) = \mathbf{0}_{I \times k}$, $\frac{Cov(\mathbf{D}, f(\mathbf{X}) - \mathbf{X}\boldsymbol{\Gamma})}{Var(\mathbf{D})} = \frac{Cov(\mathbf{D}, f(\mathbf{X}))}{Var(\mathbf{D})}$. We state the left hand side of the equality in order to emphasize the role that misspecification plays in determining bias.

¹³ For brevity's sake, we omit the description of the conditioning state in the *plim* expression, here and subsequently.

places, (ii) the degree of correlation between D and $f(\mathbf{X})$, and (iii) the degree of functional form misspecification (i.e., $f(\mathbf{X}) - \mathbf{X}\Gamma$).

Relaxing the requirement of uniqueness in order to obtain better matches. QECG studies typically force each match of a treatment place with a control place to be unique. That is, a control place can only be matched to one treatment place. In practice, it is quite common to find that the same control place is the “best” match for more than one treatment place. When this occurs, it is standard procedure to use second-, third-, and lower-ranked control places to substitute for the higher-ranked places that are already employed as matches for other treatment places. While this allows the researcher to maintain the assumption of independence across observations (i.e., $Cov[\omega_i, \omega_s] = 0$), it comes at the cost of throwing away better matches in favor of poorer matches.

$$\text{Equation (12.b) states that } Bias(\hat{\beta}) = E \left(\left| \frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \right| \bullet \right). \quad \text{An}$$

implication of this result is that poorer matches increase bias. It follows that the requirement of uniqueness inflates the bias associated with QECG estimation under imperfect matching. This immediately suggests an improvement to existing QECG methods that should result in less biased estimates of treatment effects.

Since the requirement of uniqueness results in poorer matches, QECG analyses should relax this requirement and, instead, match each treatment place with its “closest” control places, irrespective of whether these control places are matched to other treatment

places.¹⁴ Of course, if the same control places were used for more than one treatment place, correlations would be generated across observations, violating the assumption of independence. Accordingly, we develop two QECG estimators, depending on whether the degree of imperfect matching is assumed to be “negligible”.

Case One: Imperfect matching is assumed to be “negligible”. If we are willing to assume that differences in the characteristics of the matched treatment and control places

are “negligible” so that $\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \approx 0$ for $i = 1, 2, \dots, T$, then

$$\omega_i = \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \right) + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right) \approx \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right). \text{ In that case,}$$

$$(15) \quad \text{Var}[\omega_i] \approx \left(\frac{P+1}{P} \right) \sigma_\varepsilon^2 \text{ and}$$

$$(16) \quad \text{Cov}[\omega_i, \omega_s] \approx \frac{k_{is} \sigma_\varepsilon^2}{P^2},$$

where k_{is} is the number of control counties that directly overlap for treatment counties i and s , $0 \leq k_{is} \leq P$.

Define $\tilde{\beta}$ as the minimum variance QECG estimator obtained by estimating β in equation (8) when (i) each treatment place is matched with its closest control places (thus allowing the same control places to be used across observations) and (ii) imperfect matching is assumed negligible. Let $\tilde{\Sigma}$ estimate the $\text{Var}[\omega]$, where the diagonal and off-

¹⁴ Another implication of equation (12.b) is that less biased estimates will be produced when $P=1$, i.e., when each treatment place is matched to only one control place. This issue is discussed below.

diagonal elements of $Var[\omega]$ are defined by the right hand sides of equations (15) and (16), respectively. Then $\tilde{\beta}$ is simply the feasible GLS estimator of β , where

$$(17) \quad \tilde{\beta} = (\mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{i})^{-1} \mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{y}, \text{ and}$$

$$(18) \quad Var(\tilde{\beta}) = (\mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{i})^{-1}.$$

Case Two: Imperfect matching cannot be assumed to be “negligible”. Suppose

the researcher is not willing to assume $\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \approx 0$ for $i = 1, 2, \dots, T$, even

when using the best possible matches for each treatment place. If imperfect matching cannot be ignored, then biasedness in the QECG estimator is unavoidable. However, it may be possible to exploit heterogeneity in the closeness of the matches to obtain a better estimator of the treatment effect. An estimator that weights observations characterized by the closeness of matches may be able to improve on QECG methods that ignore heterogeneity in the quality of the matches.¹⁵

To implement this idea, let us assume that

¹⁵ There are, of course, many ways to exploit this heterogeneity in estimation. An alternative/complementary approach is to match each treatment to those control places that lie within a certain “closeness” range. Using a predetermined Mahalanobis distance criterion, one would allow P to vary by observation. A possibility is that $P=0$ for some treatments. Observations with no matches meeting the closeness criterion would be excluded from the analysis. One criticism of this approach is that it does not address the remaining heterogeneity in the quality of matches. A practical concern is the ad hoc selection of the “closeness” criterion.

$$(19) \quad \frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} = \left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right)^{1/2} \lambda_i ,$$

where λ_i is assumed to have a conditional distribution characterized by mean μ_λ

$$(\text{recognizing that } E \left[\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \middle| \bullet \right] \neq 0) \text{ and variance } \sigma_\lambda^2 .$$

While the specification of equation (19) is adhoc, it has at least two desirable properties. First, it implies that the mean absolute value of the error term ω_i is an increasing function of the average Mahalanobis distance of the control places. Second, in the case of perfect matches, $M_{i,p} = 0$, $p=1,2,\dots,P$, for all i , and the problem specializes to the analysis of Section 2. It follows from equation (19) that

$$(20) \quad \text{Var}[\omega_i | \bullet] = \text{Var} \left[\left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \right) + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right) \middle| \bullet \right]$$

$$= \text{Var} \left[\left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right)^{1/2} \lambda_i + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right) \middle| \bullet \right] .$$

If we further assume that

$$(21.a) \quad E[\lambda_i \varepsilon_s | \bullet] = 0 \text{ for all } i \text{ and } s, \text{ and}$$

$$(21.b) \quad E[\lambda_i \lambda_s | \bullet] = 0 \text{ for all } i \neq s$$

then

$$(22) \quad Var[\omega_i | \bullet] = \left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right) \sigma_\lambda^2 + \left(\frac{P+1}{P} \right) \sigma_\varepsilon^2$$

and

$$(23) \quad Cov[\omega_i, \omega_s | \bullet] = \frac{k_{is} \sigma_\varepsilon^2}{P^2},$$

where k_{is} is the number of control counties that directly overlap for treatment counties i and s , $0 \leq k_{is} \leq P$.

Employing the logic of GLS, define the weighting matrix $\tilde{\Sigma}$ as the estimate of $Var[\omega | \bullet]$, where the diagonal and off-diagonal elements are defined by equations (22) and (23), respectively. Note that $Var[\omega | \bullet]$ consists of unknown parameters σ_λ^2 and σ_ε^2 . To estimate these, let us define the auxiliary equation

$$(24) \quad e_i^2 = \left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right) \sigma_\lambda^2 + \left(\frac{P+1}{P} \right) \sigma_\varepsilon^2 + \gamma_i ; \text{ where}$$

$$(25) \quad e_i = y_{Ti} - \frac{\sum_{p=1}^P y_{Ci,p}}{P} - \hat{\beta},$$

γ_i is a classical error term, and $\hat{\beta}$ is the estimator of β defined in equation (10). In

other words, e_i is the residual from regressing $y_{Ti} - \frac{\sum_{p=1}^P y_{Ci,p}}{P}$ on a constant term. We

propose that σ_λ^2 and σ_ε^2 be estimated via least squares estimation of equation (24).¹⁶

The corresponding weighted QECG estimator is thus

$$(26) \quad \tilde{\tilde{\beta}} = \left(\mathbf{i}' \tilde{\tilde{\Sigma}}^{-1} \mathbf{i} \right)^{-1} \mathbf{i}' \tilde{\tilde{\Sigma}}^{-1} \mathbf{y}, \text{ where}$$

$$(27) \quad \tilde{\tilde{\Sigma}}^{-1} = \begin{bmatrix} \hat{s}_{11} & \hat{s}_{12} & \cdots & \hat{s}_{1T} \\ \hat{s}_{21} & \hat{s}_{22} & \cdots & \hat{s}_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{s}_{T1} & \hat{s}_{T2} & \cdots & \hat{s}_{TT} \end{bmatrix}^{-1};$$

$$(28) \quad \hat{s}_{ii} = \left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right) \hat{\sigma}_\lambda^2 + \left(\frac{P+1}{P} \right) \hat{\sigma}_\varepsilon^2, \quad i = 1, 2, \dots, T; \text{ and}$$

$$(29) \quad \hat{s}_{is} = \frac{k_{is} \hat{\sigma}_\varepsilon^2}{P^2}, \quad i \neq s.$$

The standard error of $\tilde{\tilde{\beta}}$ must be estimated by resampling procedures (e.g. bootstrapping), as analytic estimates of the variance cannot be calculated.¹⁷

¹⁶ For a discussion of this approach in a similar setting, see Johnston and DiNardo (1997, page 172). However, unlike the application discussed there, note that estimation of the auxiliary equation, even given the validity of the underlying assumptions, does not produce consistent estimates of σ_λ^2 and σ_ε^2 , since $\hat{\beta}$ is not a consistent estimator of β (given imperfect matching, there is no consistent estimator of β).

¹⁷ To see this, note that

It is important to acknowledge that the properties of $\tilde{\beta}$ are analytically unknown. While the logic of the estimator parallels GLS estimation, it departs from conventional GLS estimation in a number of ways. Most importantly, the existence of imperfect matching implies that QECG estimators (as well as CR estimators) will be biased. There is no avoiding this conclusion. The best that the researcher can hope to accomplish is to choose an estimation procedure that minimizes the associated bias. That being said, we note that the Monte Carlo simulation work that follows this analysis will demonstrate that weighted QECG estimators can result in a dramatic improvement in estimator performance.

Matching treatment places to multiple control places. In contrast to twins studies where each treatment place is matched with a single control place, the preceding analysis allows each treatment place to be matched with multiple control places ($P \geq 1$). In this section we demonstrate why it might be advantageous to match treatment places to more

$$E(\tilde{\beta}) = \beta + E \left[\left(\mathbf{i}' \tilde{\mathbf{Z}}^{-1} \mathbf{i} \right)^{-1} \mathbf{i}' \tilde{\mathbf{Z}}^{-1} \left[\left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right)^{1/2} \lambda_i + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right) \right] \right]. \quad \text{The expectation}$$

operator cannot be moved inside the parentheses because the elements of $\tilde{\mathbf{Z}}$ are stochastic. The corresponding variance of $\tilde{\beta}$ is given by

$$\text{Var}(\tilde{\beta}) = E \left[\left(\mathbf{i}' \tilde{\mathbf{Z}}^{-1} \mathbf{i} \right)^{-1} \mathbf{i}' \tilde{\mathbf{Z}}^{-1} \mathbf{y} - E \left[\left(\mathbf{i}' \tilde{\mathbf{Z}}^{-1} \mathbf{i} \right)^{-1} \mathbf{i}' \tilde{\mathbf{Z}}^{-1} \left[\left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right)^{1/2} \lambda_i + \left(\varepsilon_{Ti} - \frac{\sum_{p=1}^P \varepsilon_{Ci,p}}{P} \right) \right] \right] \right]^2$$

which is analytically intractable. Further, asymptotic analysis is of little value since we do not have a consistent estimator of σ_λ^2 and σ_ε^2 .

than one control place. The argument is best seen by considering the least squares estimator $\hat{\beta}$, defined in equation (10), in the special case where no control place is matched to more than one treatment place.

$$\text{As shown above, } Bias(\hat{\beta}) = E \left(\left| \frac{\sum_{p=1}^P (f(\mathbf{X}_{T_i}) - f(\mathbf{X}_{C_{i,p}}))}{P} \right| \bullet \right). \text{ If control places are}$$

matched in order of closeness, then $Bias(\hat{\beta})$ will be a strictly increasing function of P . In other words, if one is interested in minimizing bias, then each treatment place should be matched to only one control place.

Typically, however, researchers are not interested in minimizing bias, but mean-square-error (MSE), (or some function of MSE).¹⁸ MSE may be decomposed into both a variance and a bias component, so that

$$(30) \quad MSE(\hat{\beta}) = Variance(\hat{\beta}) + Bias(\hat{\beta})^2.$$

The preceding analysis has established that the bias component of $MSE(\hat{\beta})$ will increase as poorer matches are included in the analysis. However, adding more matches may decrease the variance component of $MSE(\hat{\beta})$.

¹⁸ Recall that adjusted R-squared, the Akaike Information Criterion (AIC), and the Schwarz Information Criterion (SIC) are all functions of MSE.

Equation (22) states that $Var[\omega_i | \bullet] = \left(\frac{\sum_{p=1}^P M_{i,p}}{P} \right) \sigma_\lambda^2 + \left(\frac{P+1}{P} \right) \sigma_\varepsilon^2$, where P is the

number of matches. When $P=1$, $\left(\frac{P+1}{P} \right) = 2$. When $P=3$, $\left(\frac{P+1}{P} \right) = 1.33$. And when $P=5$, $\left(\frac{P+1}{P} \right) = 1.2$. Thus, increasing P from 1 to 3 or more can potentially result in a substantial reduction in the second component of $Var[\omega_i | \bullet]$, especially when σ_ε^2 is relatively large.

When no control place is matched to more than one treatment place, the error terms ω_i are independent across observations so that

$$(31) \quad Var(\hat{\beta}) = Var\left(\beta + \frac{\mathbf{i}'\omega}{T} \mid \bullet\right) = \frac{Var(\omega_i \mid \bullet)}{T}$$

$$= \left(\frac{\sum_{p=1}^P M_{i,p}}{TP} \right) \sigma_\lambda^2 + \left(\frac{P+1}{TP} \right) \sigma_\varepsilon^2.$$

Substituting equations (12.b) and (31) into (30) yields

$$(32) \quad MSE(\hat{\beta}) = \left(\frac{\sum_{p=1}^P M_{i,p}}{TP} \right) \sigma_\lambda^2 + \left(\frac{P+1}{TP} \right) \sigma_\varepsilon^2 + E \left(\frac{\sum_{p=1}^P (f(\mathbf{X}_{Ti}) - f(\mathbf{X}_{Ci,p}))}{P} \mid \bullet \right)^2.$$

Equation (32) identifies the cost and benefit of matching treatment places to more than one control place. On the cost side, including poorer matches causes an increase in the first and third components of $MSE(\hat{\beta})$. On the benefit side, including more matches

(increasing P) lowers the second component of $MSE(\hat{\beta})$. If the additional control places are nearly as close as the best/first match, then the benefit of matching treatment places to multiple control places may outweigh the cost. In other words, while matching treatment places to more than one control place will increase the bias of the estimator $\hat{\beta}$, it may decrease the associated MSE .

Unfortunately, the analytical intractability of the QECG estimators $\tilde{\beta}$ and $\tilde{\tilde{\beta}}$ does not allow us to formally extend the preceding result to these estimators. Nevertheless, the illustrative case of $\hat{\beta}$ suggests that there may be advantages to using multiple control places. We will investigate this possibility in the Monte Carlo experiments below.

Summary. Our analysis considers three QECG estimators, each resting on a different set of assumptions. The first estimator, $\hat{\beta}$, assumes that it is possible to match each control place to only one treatment place and still achieve close to perfect matches. If the degree of match imperfection is “negligible”, so that

$$E \left(\left| \frac{\sum_{p=1}^P (f(X_{Ti}) - f(X_{Ci,p}))}{P} \right| \bullet \right) \approx 0, \text{ then the standard OLS estimators of } \beta \text{ and } \sigma_{\beta} \text{ in}$$

equation (8) will be “almost” unbiased.

The second estimator, $\tilde{\beta}$, assumes that it is possible to achieve close to perfect matches, but to do that one must allow control places to be matched to more than one treatment place. If the associated degree of match imperfection is negligible, then the GLS estimators $\tilde{\beta} = (\mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{i})^{-1} \mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{y}$ and $\hat{\sigma}_{\tilde{\beta}} = \sqrt{(\mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{i})^{-1}}$ (cf. equations (15) through

(18)) will be “almost” unbiased, where the weighting matrix $\tilde{\Sigma}$ accounts for the fact that the same control places appear in more than one observation.

The third estimator applies to those applications where the researcher is not willing to assume that the degree of match imperfection is negligible, even when control places are allowed to be matched to more than one treatment place. In these instances we propose the estimator $\tilde{\beta} = \left(\mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{i} \right)^{-1} \mathbf{i}' \tilde{\Sigma}^{-1} \mathbf{y}$, where $\tilde{\Sigma}$ is a weighting matrix that weights “good” matches more heavily than “poor” matches (while also accounting for the appearance of control places in multiple observations). Unfortunately, the analytic properties of this estimator are unknown.

This analysis has formulated all three estimators to allow treatment places to be matched to more than control place (i.e., $P > I$). While including poorer quality matches is likely to increase bias, it may result in decreased MSE. Whether this potentiality is likely to be realized in practice is an unanswered question.

CR methods (represented by OLS estimation of equation (2)) will produce asymptotically unbiased estimates of policy impacts when either $\text{plim} \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) = 0$ or

$f(\mathbf{X}) = \mathbf{X}\boldsymbol{\Gamma}$. If $\text{plim} \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) = 0$, it follows from equation (13.c) that

$$(33.a) \quad \text{Bias}(b) = \left(\text{plim} \left(\frac{\mathbf{D}_*' \mathbf{D}_*}{N} \right) \right)^{-1} \text{plim} \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) = 0.$$

Likewise, if $f(\mathbf{X}) = \mathbf{X}\boldsymbol{\Gamma}$, then

$$\begin{aligned}
(33.b) \quad Bias(b) &= \left(plim \left(\frac{\mathbf{D}_*' \mathbf{D}_*}{N} \right) \right)^{-1} plim \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) \\
&= \left(plim \left(\frac{\mathbf{D}_*' \mathbf{D}_*}{N} \right) \right)^{-1} plim \left(\frac{\mathbf{D}'(\mathbf{M}_* \mathbf{X}) \boldsymbol{\Gamma}}{N} \right) = 0.
\end{aligned}$$

Thus, both the nonrandom adoption of treatments (defined in this context as

$$plim \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) \neq 0) \text{ and misspecification } (f(\mathbf{X}) \neq \mathbf{X}\boldsymbol{\Gamma}) \text{ are necessary for CR methods}$$

to produce (asymptotically) biased estimates.

This allows us to identify eight scenarios for comparing QECG and CR estimators, depending on whether (i) the process determining treatment places is random or nonrandom, (ii) treatment places are matched perfectly or imperfectly, and (iii) the CR equations are correctly or incorrectly specified. These eight scenarios are represented in Table 1.

When treatment places are either randomly determined (defined as $Cov(D, f(\mathbf{X})) = 0$) or there is perfect matching ($\mathbf{X}_{Ti} = \mathbf{X}_{Ci,p}$ for all i, p), QECG estimators will be unbiased (Scenarios (1)-(6)). If both $Cov(D, f(\mathbf{X})) \neq 0$ and $\mathbf{X}_{Ti} \neq \mathbf{X}_{Ci,p}$ for all i, p , QECG estimators will be biased (Scenarios (7) and (8)).

Alternatively, when treatment places are either randomly determined (defined in this context as $plim \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) = 0$) or the CR equation is correctly specified

($f(\mathbf{X}) = \mathbf{X}\boldsymbol{\Gamma}$), the CR estimator will be unbiased (Scenarios (1)-(5) and (7)).¹⁹ If both

$plim\left(\frac{\mathbf{D}_*'\mathbf{f}(\mathbf{X})}{N}\right) \neq 0$ and $\mathbf{f}(\mathbf{X}) \neq \mathbf{X}\boldsymbol{\Gamma}$, the CR estimator will be biased (Scenarios (6)

and (8)).

Unfortunately, many of the key questions—particularly concerning the relative performance of the QECG and CR estimators—are analytically intractable. As a result, the next section continues our study using Monte Carlo experiments. The eight scenarios of Table 1 provide the framework for organizing our experiments. Of these, scenario (8)—described by “nonrandom determination of treatment places/imperfect matching/misspecification in the CR equation”—represents the empirical environment most likely to be encountered in practice. We pay particular attention to this scenario in the subsequent analysis.

IV. MONTE CARLO ANALYSIS OF CR AND QECG ESTIMATORS

This section employs Monte Carlo experiments to study three estimators of the treatment effect β in equation (2): (i) the CR estimator b , (ii) the QECG estimator $\tilde{\beta}$ (defined in equation (17), and the QECG estimator $\tilde{\tilde{\beta}}$ (defined in equation (26)).²⁰ The

¹⁹ We note that when $Cov(\mathbf{D}, \mathbf{X}) = \mathbf{0}_{I \times K}$, $plim\left(\frac{\mathbf{D}_*'\mathbf{f}(\mathbf{X})}{N}\right) = Cov(\mathbf{D}, \mathbf{f}(\mathbf{X}))$, and the two definitions of random determination of treatment places coincide.

²⁰ We do not study the estimator $\hat{\beta}$ (defined in equation (10), since that would require a procedure for insuring that each control place was matched to only one treatment place. As noted above, a number of different algorithms are used in the literature and the algorithms tend to be researcher-specific. One can think of the QECG estimator $\tilde{\beta}$ as

QECG estimators $\tilde{\beta}$ and $\tilde{\tilde{\beta}}$ correspond to the cases where imperfect matching is assumed—perhaps incorrectly—to be “negligible” and “not negligible”, respectively.

The experiments consist of variations on the following five components:

1. **(The Distribution of Place Characteristics, X_i .)** Characteristic k for place i is generated by the process $X_{ki} = x_{ki} + \theta \varepsilon_{X,ki}$, $k=1,2,\dots,K$, where x_{ki} is distributed uniformly between 0 and 10, and $\varepsilon_{X,ki} \sim N(0,1)$. In most of the experiments, $K=2$, so that places are described by two characteristics.

The parameter θ regulates the similarity between treatment and control places. The data generation algorithm is constructed so that when $\theta=0$, there is at least one control place that has identical X_i values for each treatment place. In other words, the experiment $\theta=0$ corresponds to the case of “perfect matching.” The larger θ , the larger will be the spread of X_{ki} values within a set of related treatment-control places (i.e., treatment and control places generated from the same initial x_{ki} values) and, ceteris paribus, the greater the dissimilarity between treatments and controls.

2. **(The Function $f(X_i)$.)** The function $f(X_i)$ is generally specified as

$$f(X_i) = \sum_{k=1}^K X_{ki} + \sum_{k=1}^K (X_{ki} - 5)^2.$$

Noteworthy here is that $f(X_i)$ is a nonlinear function

in X_i , and that the quadratic component is specified to be orthogonal to the linear component.

bounding these estimators in terms of minimum bias since it always uses the best matches.

3. **(The Determination of Treatment and Control Places.)** Treatment and control places are identified by the variable D , which takes values of 1 and 0. The variable D_i is generated by a two-stage process. In the first stage, the latent variable d_i is determined

$$\text{according to } d_i = \begin{cases} \tau + \text{step}, & \text{if } \tau + \text{step} \leq \sum_{k=1}^K (X_{ki} - 5)^2 \\ \sum_{k=1}^K (X_{ki} - 5)^2, & \text{if } \tau - \text{step} \leq \sum_{k=1}^K (X_{ki} - 5)^2 < \tau + \text{step} \\ \tau - \text{step}, & \text{if } \sum_{k=1}^K (X_{ki} - 5)^2 < \tau - \text{step} \end{cases}$$

In the second stage, D_i is determined according to $D_i = \begin{cases} 1, & \text{if } d_i + \varepsilon_{D,i} \geq \tau \\ 0, & \text{if } d_i + \varepsilon_{D,i} < \tau \end{cases}$, where

$\varepsilon_{D,i} \sim N(0,1)$, and τ is a parameter that determines the proportion of observations that are treatment places. The specification of D_i is chosen so that $\text{Cov}(D, f(\mathbf{X})) \neq 0$ and $\text{Cov}(D, \mathbf{X}) = \mathbf{0}_{1 \times K}$.

The parameter “step” regulates $\text{Cov}(D, f(\mathbf{X}))$, and hence bias (cf. equation (14)). Larger (smaller) values of “step” serve to increase (decrease) bias. “step” is generally set equal to 2, except in Table 5b where it is decreased to show the robustness of the Set 3 experiments to changes in $\text{Cov}(D, f(\mathbf{X}))$.

4. **(The True Model Relating Outcomes to Treatment).** The true, unobserved model relating y_i to D_i is given by $y_i = 0 + 0 \cdot D_i + 1 \cdot f(\mathbf{X}_i) + \varepsilon_{y,i}$, where $\varepsilon_{y,i} \sim N(0,36)$. This specification states that the true policy impact is zero. Thus, any estimated relationship between the outcome variable y_i and the treatment variable D_i is spurious and constitutes bias.

5. **(The Misspecified CR Model).** The CR model is generally specified to be

$y_i = \alpha + \beta D_i + \sum_{k=1}^K \gamma_k X_{ki} + \text{error term}$, where y_i , D_i , and X_i are observed, and α , β , and $\gamma_1, \dots, \gamma_K$ are parameters to be estimated. Note that this model is misspecified in that it omits the nonlinear terms in $f(X_i)$.

The stochastic nature of the data generation process, combined with the requirement that every treatment place have at least one perfectly matched control place when $\theta = 0$, implies that the number of observations in any given experimental sample cannot be exactly controlled. However, parameters can be adjusted to inexactly control the number and composition of sample observations. We set parameters to produce CR samples of approximately 395 observations (i.e., $N \approx 395$), composed of approximately 65 treatment observations ($T \approx 65$) and 330 control observations.²¹ Each QECG sample is formed from a corresponding CR sample by matching a subset of the control observations to the full set of treatment observations appearing in that CR sample. Each “experiment” consists of 1000 replications (or 1000 samples).

Table 2 describes the five sets of Monte Carlo experiments performed in this study. The first three sets of experiments are designed to illustrate the roles of (i) misspecification in the CR model and (ii) nonlinearity in $f(X)$ with respect to the absolute and relative performances of the CR and QECG estimators. The fourth set of experiments studies the QECG estimator $\tilde{\beta}$ to determine whether the benefits of

²¹ These numbers are representative of a QECG analysis of a state-wide policy using county-level data. The average U.S. state has 63 counties and the average region, as defined by the Bureau of Economic Analysis, has about 395 counties.

matching more than one control place to each treatment place outweigh the costs. The fifth set of experiments examines the implications of including irrelevant variables for the absolute and relative performance of the CR and QECG estimators.²² All of the experiments are programmed in SAS/IML. The associated computer programs may be downloaded from the internet.²³

Set 1 experiments. The experiments in Set 1 correspond to Scenarios (1) and (3) in Table 1. The results of the Monte Carlo experiments are reported in Table 3. Note that θ increases as one moves down the table. This causes poorer matches between treatment and control places, as evidenced by larger “Mean Mahalanobis” values. The experiment $(\theta = 0, \tau = 45)$ represents the case of perfect matching. Subsequent experiments in Table 3 represent imperfect matching.

The Set 1 experiments study the effect on the QECG and CR estimators of increasing dissimilarity between treatment and control places when (i) there is no misspecification in the estimated CR equation and (ii) $f(X)$ is linear. The linearity of

²² We do not investigate the effect of omitting relevant variables from X . Unlike the case of adding irrelevant variables, this form of misspecification is not easily incorporated within the framework of misspecification set out in Section 2. Nevertheless, we did attempt to study the issue using Monte Carlo experiments. The results are not easily summarized. Sometimes the CR estimator outperformed the QECG estimators, and sometimes the opposite was true. Patterns were difficult to discern. Accordingly, we leave this as a topic for future research.

²³ Programs may be downloaded as WORD documents from “<http://faculty-staff.ou.edu/R/William.R.Reed-1/Papers>”.

$f(\mathbf{X})$ implies that treatment places are randomly determined.²⁴ As a result, both QECG and CR estimators are unbiased.

The experiments confirm the unbiasedness of the estimators. The estimates associated with the CR estimator b are centered on 0, the true value of β . The QECG estimates associated with $\tilde{\beta}$ and $\tilde{\tilde{\beta}}$ are also centered on 0. While each of the QECG and CR estimators is unbiased, the CR estimates are generally superior. They are closer to zero and have smaller MSE's than the corresponding QECG estimates. Finally, there appears to be little difference between the performance of the QECG estimators $\tilde{\beta}$ and $\tilde{\tilde{\beta}}$ in this set of experiments.

Set 2 experiments. The experiments in Set 2 correspond to Scenarios (5) and (7) in Table 1. Table 4 reports the results of the corresponding experiments. Once again, the experiment $(\theta = 0, \tau = 45)$ represents the case of perfect matching, with subsequent experiments representing imperfect matching.

This second set of experiments is similar to the first except that $f(\mathbf{X})$ is made to be nonlinear. As a result, treatment places are nonrandomly determined. This change has no effect on the bias of the CR estimates given that the CR equation is not misspecified. However, it does adversely impact the QECG estimates.

²⁴ The combination of (i) $f(\mathbf{X}) = \sum_{k=1}^2 X_k$, (ii) D a function of $\sum_{k=1}^2 (X_k - 5)^2$, and (iii) $Cov(D, \mathbf{X}) = \mathbf{0}_{I \times K}$, implies that $plim \left(\frac{\mathbf{D}' f(\mathbf{X})}{N} \right) = Cov(D, f(\mathbf{X})) = 0$.

The nonlinearity of $f(\mathbf{X})$ implies that $Cov(D, f(\mathbf{X})) > 0$, so that treatment places will generally be characterized by larger values of $f(\mathbf{X})$ than control places. The outcome variable y is also positively correlated with $f(\mathbf{X})$ (since $y_i = 0 + 0 \cdot D_i + 1 \cdot f(\mathbf{X}_i) + \varepsilon_{y,i}$). This generates a positive correlation between D and y . As long as treatment places are perfectly matched, the QECG estimates will be unbiased. However, we expect to see evidence of positive bias when matching is imperfect (cf. equation (12.b)).

The Set 2 experiments again demonstrate the unbiasedness of the CR estimator b . In contrast, as the degree of imperfect matching increases (as evidenced by larger “Mean Mahalanobis” values), one observes that the QECG estimates become increasingly positive, displaying greater bias and MSE values. These results are consistent with the analytical results we derived earlier, summarized in Table 1 under Scenario (7). When (i) places are nonrandomly determined, (ii) there is no misspecification in the CR equation, and (iii) treatment and control places are imperfectly matched, the CR estimator is superior to the QECG estimators.

Interestingly, the QECG estimator $\tilde{\beta}$ --which weights observations by the closeness of matches--performs substantially better than $\tilde{\beta}$ --which does not. Not only does it have smaller bias in every experiment, but it is also characterized by smaller MSE values. On the basis of these and other experiments, we find that there appear to be substantial benefits to weighting observations by the closeness of matches.

When the experimental environment is such that QECG estimators are biased—such as in Table 4--weighting by the closeness of matches reduces the bias. When the

experimental environment is such that QECG estimators are unbiased—such as in Table 3--weighting appears to have little cost in terms of MSE. In other words, imposing the assumption that imperfect matching is “negligible” has little to recommend it. It can cause substantial costs in estimator performance in estimation environments where imperfect matching is problematic, without generating compensating advantages when it is not. This finding should be of interest to researchers desiring more reliable estimates of policy impacts using QECG analysis.

Set 3 experiments. The experiments in Set 3 correspond to Scenarios (6) and (8) in Table 1. These experiments represent the main focus of this study. Of particular interest is Scenario 8 (“nonrandom determination of treatment and control places/misspecification in the CR equation/imperfect matching”), since this characterizes the empirical setting most likely to be encountered in practice.

The preceding theoretical analysis has demonstrated that QECG methods will be superior to CR methods when (i) the treatment variable is endogenous but all relevant variables are included in the analysis (so that the treatment variable is uncorrelated with the error term in the true model), (ii) the functional relationship between y and the explanatory variables is nonlinear and unknown (so that the CR model is misspecified), and (iii) each treatment place can be matched to a control place having identical characteristics (i.e., the case of “perfect matching”). When there is imperfect matching, the advantages of QECG methods relative to CR methods are unclear. However, our theoretical analysis suggests that (i) the closeness of matches between treatment and control places and (ii) the degree of correlation between D and $f(X)$ are both likely to be important factors.

The Set 3 experiments are designed to empirically investigate these issues. The first group of experiments that we perform in this set is made to be identical to those reported in Table 4 except that the CR equations are now misspecified. In particular, the CR equations omit the nonlinear component of $f(\mathbf{X})$. This induces a spurious, positive correlation between y and D , even after controlling for the other variables in the CR equation. Since the QECG estimates will be the same as before, the major interest in these experiments is the relative performance of the CR and QECG estimators.

Table 5A reports results from this first group of experiments. The first row in the table (experiment $(\theta = 0, \tau = 42)$) allows a comparison of CR estimates with QECG estimates under perfect matching. This experiment represents the “textbook case” that highlights the advantages of QECG estimators (cf. Section 2 above). Perfect matching overcomes the problem of misspecification when the functional form of $f(\mathbf{X})$ is unknown. Monte Carlo study of this experiment empirically demonstrates the superiority of QECG analysis. In the presence of (i) misspecification in the CR model and (ii) nonlinearity in $f(\mathbf{X})$, the CR estimator b has a bias of 5.5083 and a MSE of 30.9679. In contrast, the QECG estimator $\tilde{\beta}$ has a bias of -0.0770 and a MSE of 1.2513.

The subsequent rows in Table 5A investigate the relative performances of the CR and QECG estimators as the dissimilarity between treatment and control observations increase, generated by increasing values of θ . As the X ’s spread out, the quadratic component of the error term becomes larger, increasing the correlation between D and the error term in the CR equation. *Ceteris paribus*, this increases the bias of the CR estimator b . However, it also gets us out of the world of perfect matching and into the world of

imperfect matching, where the advantages of the QECG estimators are theoretically unclear.

As discussed in the context of the previous set of experiments, as one moves down the rows of Table 5A and the matches between treatments and controls becomes poorer, the QECG estimators perform worse in absolute terms. However, *in relative terms*, the QECG estimates strictly dominate the CR estimates throughout the range of experiments. For example, for the experiment $(\theta = 1.0, \tau = 45)$, the CR estimate displays a bias of 13.3081 and a MSE of 177.7892. In contrast the QECG estimator $\tilde{\beta}$ has a bias of 2.2747 and a MSE of 7.0200. For all of the experiments under imperfect matching, the MSE's of $\tilde{\beta}$ represent a dramatic improvement compared to that of the CR estimator b , generally being smaller by a factor of 10 or more. This experimental finding demonstrates what the preceding theoretical analysis could only suggest. Namely, that QECG estimators can produce substantial improvements over CR estimates, even when matching is imperfect.

The results in Table 5A are representative of a wider set of experiments that speak to the robustness of the superiority of the QECG estimators under the regime of (i) nonrandom determination of treatment and control places, (ii) misspecification in the CR equation, and (iii) imperfect matching. A sampling of this wider set of experiments is given in Table 5B, which repeats the experiments of Table 5A except that the “*step*” parameter is changed from “2” to “1” (cf. “The Determination of Treatment and Control Places” above). This causes greater noise in the treatment variable D , reducing the correlation between D and $f(X)$. This should decrease the biases associated with both the CR and QECG estimators.

The second and third columns of Table 5B compare the correlation between D and $f(X)$ for each of the experiments when “step” equals 2 and 1, respectively. The correlation between D and $f(X)$ is smaller for each experiment given the smaller “step” value. Correspondingly, both bias and MSE are lower in each experiment for the CR and QECG estimators, as is evident by a comparison of the respective values in Tables 5B and 5A. Most important for our purposes, however, is that the relative performance of the QECG estimates continues to strictly dominate the CR estimates throughout the range of experiments.

It is important to remember that this robustness in the superior performance of the QECG estimators is conditional on the general experimental environment used in this study, and that we designed this environment to recreate those circumstances in which QECG analysis is likely to outperform CR methods. A comparison of QECG and CR estimators in other experimental settings, say with different functional forms of $f(X)$ and/or different types of misspecification in the CR model would yield different results (as suggested by the Set 1 and Set 2 experiments).

Nevertheless, these findings are valuable to researchers interested in estimating state and local policy impacts. They establish the potential advantages of QECG methods over CR methods in realistic estimation environments. Further, they make it clear why one might expect QECG methods to produce better estimates of policy impacts in the presence of endogeneity and misspecification.

Set 4 experiments. The Set 4 experiments use the same data as the Set 3 experiments, except now the QECG estimators are calculated allowing for treatment places to be matched to multiple control places. Previous experiments all employ paired

matching in the QECG analyses. In each case, a treatment place is matched to its closest control place. The Set 4 experiments investigate whether there are any advantages to matching treatment places to multiple control places.

We can summarize our findings by saying that we found little evidence that the benefits of more control places outweigh the costs. Representative of this set of experiments are the results in Table 6. In general, including more control places in QECG analyses increases both bias and MSE.

There are exceptions, however. Experiment $(\theta = 0.5, \tau = 42)$ (cf. the second row in Table 6) provides an example where bias and MSE decrease as the number of control places is increased from 1 to 5. Even more interesting is the experiment $(\theta = 0, \tau = 42)$ (cf. Row 1 of Table 6). When P increases from 3 to 5, bias increases (from 0.2694 to 0.3530), but MSE decreases (from 0.9541 to 0.8704). This provides an empirical confirmation of the theoretical possibility raised in Section 3 that adding more control places, while increasing bias, could decrease Variance sufficiently to cause an overall decrease in MSE (cf. equation (32)).

Typically though, our Monte Carlo studies find that when a benefit is observed from increasing P , either in decreased bias or MSE, the benefit is relatively small measured in percentage terms. More commonly, a larger value of P is associated with both larger bias and larger MSE. Further, the associated percentage increase in bias and MSE tends to be relatively large. For example, the average percentage decrease in MSE from adding more control places (i.e., moving one column to the right), conditional on a decrease in MSE, is 6.9 percent in Table 6. The average percentage increase, conditional on there being an increase in MSE, is 40.7 percent. Based on these findings, we conclude

that the practice of matching treatment places to one control place ($P=1$) is preferable to matching to multiple control places ($P>1$).

Set 5 experiments. The last set of experiments studies the effect of including irrelevant variables on the relative performance of the CR and QECG estimators. In this set of experiments, the true X consists of X_1 and X_2 . The parameters θ and τ are fixed at $(\theta = 1.0, \tau = 45)$ for all the experiments in the set.²⁵ The individual experiments differ only in that they include additional (irrelevant) X 's in X in both the CR and QECG estimation, where the irrelevant X 's are generated from the same individual distributions as X_1 and X_2 and are uncorrelated both with the relevant X 's and each other. The experiments in this set correspond to Scenario 8 in Table 1.

Table 7 reports the results. As expected, the inclusion of irrelevant variables has no effect on the CR estimates. In contrast, adding irrelevant variables increases both bias and MSE for the QECG estimators. The QECG matching process is designed to pick control observations that minimize the Mahalanobis distance for all the variables in X , not just the relevant variables X_1 and X_2 . The result of matching on a larger set of X 's is poorer matches in the relevant variables. This is evidenced in the table by the increasing “Mean Mahalanobis” values as additional irrelevant variables are added. *Ceteris paribus*, worse matches generally result in greater bias and greater MSE for the QECG estimators. However, the QECG estimators still perform better than the CR estimator even when the number of irrelevant variables becomes very large. The last row of Table 7 reports that

²⁵ The experiment in the first row of Table 7 is essentially the same experiment reported in row 3 of Table 5A. The results differ because the program used to generate the data in Table 7 imposes different seed values on the process that generates the random numbers used to create the data.

bias and MSE values are lower for the QECG estimators when a total of 50 variables are included in X .

V. CONCLUSION

This study examines the efficacy of using Quasi-Experimental Control Group (QECG) methods to estimate policy impacts. The analysis employs a combination of analytical and experimental techniques. We make three main contributions to the QECG literature. First, we investigate the consequences of imperfect matching for the properties of QECG methods, a concern that has been mostly overlooked in empirical QECG studies. Second, we develop an estimator that generalizes the QECG approach by (1) allowing control places to be matched to more than one treatment place (thus ensuring that the “best” match is always used), and (2) weighting observations by the closeness of the match. In so doing, we address common concerns regarding the optimal way to implement QECG methods. Finally, we use Monte Carlo analysis to study the relative performance of QECG and CR estimators.

Our results demonstrate that QECG estimators can outperform the conventional regression (CR) estimator when the adoption of policy is endogenous (nonrandom), the relationship between outcomes and policies is nonlinear, and CR equations do not correctly specify the nonlinear form of the relationship. In the case of perfect matching, QECG methods produce unbiased estimates. In the case of imperfect matching, QECG estimators are no longer unbiased. However, QECG analyses can be substantially improved by adopting an estimation procedure that weights observations according to the closeness of matches.

We conclude that QECG methods provide a viable alternative to CR methods for researchers interested in estimating policy impacts. In addition, we offer a QECG estimation procedure that both simplifies the matching procedure (by eliminating the need to find unique “best” matches for each treatment) and also provides better estimates compared with previous QECG methods. Furthermore, we hope that by addressing many implementation concerns, researchers will give QECG methods more consideration when performing place-based policy analysis.

An important caveat is that many of our empirical results are generated in an experimental environment designed to highlight the advantages of QECG analysis. Accordingly, it would be useful to compare QECG with CR methods in other experimental settings, incorporating other forms of nonlinearities and other types of misspecifications. Furthermore, there may be other QECG estimators that do a better job of exploiting heterogeneities in matches between treatment and control observations than the estimator we propose here. These are topics for future research.

REFERENCES

- Aleseved, Mostafa, Terance J. Rephann, and Andrew M. Isserman. "The Local Effects of Large Dam Reservoirs: U.S. Experience, 1975-1995." Review of Urban and Regional Development Studies 10 (Autumn 1998): 91-108.
- Bartik, Timothy J. Who Benefits from State and Local Economic Development Policies? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.
- Bartik, Timothy J. "Discussion." New England Economic Review: Proceedings of a Symposium on The Effects of State and Local Public Policies on Economic Development (March/April 1997): 67-71.
- Bohm, Peter and Hans Lind. "Policy Evaluation Quality: A Quasi-experimental Study of Regional Employment Subsidies in Sweden." Regional Science and Urban Economics 23 (1993): 51-65.
- Broder, Josef M., Teresa D. Taylor, and Kevin T. McNamara. "Quasi-Experimental Designs for Measuring Impacts of Developmental Highways in Rural Areas." Southern Journal of Agricultural Economics (July 1992): 199-207.
- Greenburg, Michael, Andrew M. Isserman, Donald Drueckenberg, Karen Lowrie, Henry Mayer, Darien Simon, David Sorenson. "Socioeconomic Impacts of US Nuclear Weapons Facilities." Applied Geography 18 (1998): 101-116.
- Isserman, Andrew M. and Paul Beaumont. "Quasi-Experimental Control Group Methods for the Evaluation of Regional Economic Development Policy." Socio-Economic Planning Sciences 23 (1989): 39-53.
- Isserman, Andrew M. and John D. Merrifield. "The Use of Control Groups in Evaluating Regional Economic Policy." Regional Science and Urban Economics 12 (1982): 43-58.
- . "Quasi-Experimental Control Group Methods for Regional Analysis: An Application to an Energy Boomtown and Growth Pole Theory." Economic Geography 63 (January 1987): 3-19.
- Isserman, Andrew M., and Terance Rephann. "The Economic Effects of the Appalachian Regional Commission." Journal of the American Planning Association 61 (Summer 1995): 345-364.
- Johnston, Jack and John DiNardo. Econometric Methods, Fourth Edition. New York: The McGraw-Hill Companies, Inc., 1997.

- Mueser, Peter, Kenneth R. Troske and Alexey Gorislavsky. "Using State Administrative Data to Measure Program Performance." Mimeo, University of Missouri-Columbia, 1999.
- Rephann, Terance J., Margaret Dalton, Anthony Stair and Andrew M. Isserman. "Casino Gambling as an Economic Development Strategy." Tourism Economics 3 (1997) 161-183.
- Rephann, Terance J., and Andrew M. Isserman. "New Highways as Economic Development Tools: An Evaluation Using Quasi-Experimental Matching Methods." Regional Science and Urban Economics 24 (1994): 723-751.
- Rosenbaum, Paul R. "Optimal Matching for Observational Studies." Journal of the American Statistical Association 84 (December 1989): 1024-1032.
- Wasylenko, Michael. "Taxation and Economic Development: The State of the Economic Literature." New England Economic Review (March-April, 1997): 37-52.
- Wojan, Timothy, and T. M. Bailey. "The Statistical Power Argument for Quasi-Experimental Matched Pair Methods." Paper presented at the Meetings of the Southern Regional Science Association, Savannah, GA, April 2-4, 1998.

TABLE 1
The Relationship Between Biasedness of the QECG and CR Estimators and
(i) Endogeneity, (ii) Matching, and (iii) Misspecification

<i>Matching Process</i>	<i>Misspecification Status</i>	<i>Process Determining Treatment Places</i>	
		<i>Random^a</i> $\left(\begin{array}{l} \text{Cov}(D, f(X)) = 0 \\ \text{plim} \left(\frac{D_*' f(X)}{N} \right) = 0 \end{array} \right)$	<i>Nonrandom^a</i> $\left(\begin{array}{l} \text{Cov}(D, f(X)) \neq 0 \\ \text{plim} \left(\frac{D_*' f(X)}{N} \right) \neq 0 \end{array} \right)$
<i>Perfect Matching</i> $(X_{Ti} = X_{Ci,p} \text{ for all } i, p)$	<i>No Misspecification</i> $(f(X) = X\Gamma)$	(1) QECG: unbiased CR: unbiased ^b	(5) QECG: unbiased CR: unbiased ^b
	<i>Misspecification</i> $(f(X) \neq X\Gamma)$	(2) QECG: unbiased CR: unbiased ^b	(6) QECG: unbiased CR: biased ^b
<i>Imperfect Matching</i> $(X_{Ti} \neq X_{Ci,p} \text{ for all } i, p)$	<i>No Misspecification</i> $(f(X) = X\Gamma)$	(3) QECG: unbiased CR: unbiased ^b	(7) QECG: biased CR: unbiased ^b
	<i>Misspecification</i> $(f(X) \neq X\Gamma)$	(4) QECG: unbiased CR: unbiased ^b	(8) QECG: biased CR: biased ^b

^a The relevant conditions for the biasedness of the QECG and CR estimators are, respectively, $Cov(D, f(\mathbf{X})) = 0 / Cov(D, f(\mathbf{X})) \neq 0$ and $plim \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) = 0 / plim \left(\frac{\mathbf{D}_*' f(\mathbf{X})}{N} \right) \neq 0$, where \mathbf{D}_* is the vector of residuals from regressing D on \mathbf{X} . Note that when $Cov(D, \mathbf{X}) = \mathbf{0}_{I \times K}$, the two conditions coincide.

^b With respect to the CR estimator, the property of “unbiased/biased” refers to asymptotic bias.

TABLE 2
Description of Monte Carlo Experiments

SET	DESCRIPTION	TRUE MODEL: y_i / D_i	ESTIMATED MODELS: CR / QECG
1	<i>Studies the effect on CR and QECG estimators of dissimilarity between treatment and control observations when (i) the CR model is not misspecified and (ii) $f(\mathbf{X})$ is linear.</i>	$y = 0 + 0 \cdot D + \sum_{k=1}^2 1 \cdot X_k$ $D = g\left(\sum_{k=1}^2 (X_k - 5)^2\right)$	<u>CR Specification:</u> $y = \alpha + \beta \cdot D + \sum_{k=1}^2 \gamma_k \cdot X_k$ <u>QECG Matching:</u> Matches on (X_1, X_2) ; $P=1$
2	<i>Studies the effect on CR and QECG estimators of dissimilarity between treatment and control observations when (i) the CR model is not misspecified and (ii) $f(\mathbf{X})$ is nonlinear.</i>	$y = 0 + 0 \cdot D + \sum_{k=1}^2 1 \cdot X_k + \sum_{k=1}^2 1 \cdot (X_k - 5)^2$ $D = g\left(\sum_{k=1}^2 (X_k - 5)^2\right)$	<u>CR Specification:</u> $y = \alpha + \beta \cdot D + \sum_{k=1}^2 \gamma_k \cdot X_k + \sum_{k=1}^K \delta_k (X_k - 5)^2$ <u>QECG Matching:</u> Matches on (X_1, X_2) ; $P=1$
3	<i>Studies the effect on CR and QECG estimators of dissimilarity between treatment and control observations when (i) the CR model is misspecified and (ii) $f(\mathbf{X})$ is nonlinear.</i>	$y = 0 + 0 \cdot D + \sum_{k=1}^2 1 \cdot X_k + \sum_{k=1}^2 1 \cdot (X_k - 5)^2$ $D = g\left(\sum_{k=1}^2 (X_k - 5)^2\right)$	<u>CR Specification:</u> $y = \alpha + \beta \cdot D + \sum_{k=1}^2 \gamma_k \cdot X_k$ <u>QECG Matching:</u> Matches on (X_1, X_2) ; $P=1$
4	<i>Studies the effect on the QECG estimator $\tilde{\beta}$ when multiple control places are matched to each treatment place.</i>	$y = 0 + 0 \cdot D + \sum_{k=1}^2 1 \cdot X_k + \sum_{k=1}^2 1 \cdot (X_k - 5)^2$ $D = g\left(\sum_{k=1}^2 (X_k - 5)^2\right)$	<u>CR Specification:</u> $y = \alpha + \beta \cdot D + \sum_{k=1}^2 \gamma_k \cdot X_k$ <u>QECG Matching:</u> Matches on (X_1, X_2) ; $P \geq 1$

TABLE 2 (continued)
Description of Monte Carlo Experiments

<i>SET</i>	<i>DESCRIPTION</i>	<i>TRUE MODEL:</i> y_i / D_i	<i>ESTIMATED MODELS:</i> <i>CR / QECG</i>
5	Studies the effect on CR and QECG estimators of including irrelevant variables.	$y = 0 + 0 \cdot D + \sum_{k=1}^2 1 \cdot X_k + \sum_{k=1}^2 1 \cdot (X_k - 5)^2$ $D = g\left(\sum_{k=1}^2 (X_k - 5)^2\right)$	<p><u>CR Specification:</u></p> $y = \alpha + \beta \cdot D + \sum_{k=1}^2 \gamma_k \cdot X_k$ $y = \alpha + \beta \cdot D + \sum_{k=1}^3 \gamma_k \cdot X_k$ \vdots $y = \alpha + \beta \cdot D + \sum_{k=1}^{10} \gamma_k \cdot X_k$ <p><u>QECG Matching:</u> Matches on (X_1, X_2), $(X_1, X_2, X_3), \dots, (X_1, X_2, \dots, X_{10})$; $P=1$</p>

TABLE 3
The Effect On CR And QECG Estimators Of Increasing Dissimilarity Between Treatment And Control Observations
When There Is No Misspecification In The CR Model And $f(X)$ Is Linear

<i>Experiment</i>	<i>Mean Mahalanobis</i>	<i>N / T</i>	<i>BIAS</i>			<i>MSE</i>		
			<i>b</i>	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$	<i>b</i>	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$
$(\theta = 0, \tau = 45)$	0.0000	397 / 54	-0.0450	-0.0640	NA ^a	0.7729	1.3152	NA ^a
$(\theta = 0.5, \tau = 45)$	0.0303	396 / 60	0.0112	-0.0126	-0.0060	0.7287	1.2794	1.2963
$(\theta = 1.0, \tau = 48)$	0.0565	393 / 63	-0.0370	-0.0358	-0.0154	0.7337	1.3208	1.3183
$(\theta = 1.5, \tau = 48)$	0.1177	398 / 67	-0.0259	0.0699	0.0484	0.6224	1.3456	1.3451
$(\theta = 2.0, \tau = 48)$	0.3669	393 / 74	-0.0352	-0.1363	-0.1337	0.5894	1.5796	1.5790
$(\theta = 2.5, \tau = 70)$	0.3243	395 / 62	0.0206	0.1462	0.1562	0.6685	1.7616	1.7678
$(\theta = 3.0, \tau = 75)$	0.4782	393 / 74	-0.0043	-0.1281	-0.1275	0.6415	1.8963	1.9011

NOTE: This set of experiments corresponds to the “Set 1” experiments described in Table 2. “ θ ” is a parameter that adjusts the closeness of treatment and control observations. The experiment $\theta = 0$ represents the case of perfect matching. “ τ ” is a parameter that regulates the proportion of observations that are treatment observations. “Mean Mahalanobis” measures the average Mahalanobis distance between each treatment place and its matched control place (cf. equation (7) in the text). “N/T” reports the number of observations used in each replication of the CR and QECG estimations, respectively. “ b ” is the CR estimator; and $\tilde{\beta}(P=1)$ and $\tilde{\tilde{\beta}}(P=1)$ are the QECG estimators defined in equations (17) and (26) when each treatment place is matched to only one control place (i.e., $P=1$). Table values are the mean values associated with 1000 replications of the respective experiment.

^a The QECG estimator $\tilde{\tilde{\beta}}$ is not defined in the case of perfect matching.

TABLE 4
The Effect On CR And QECG Estimators Of Increasing Dissimilarity Between Treatment And Control Observations
When There Is No Misspecification In The CR Model And $f(X)$ Is Nonlinear

<i>Experiment</i>	<i>Mean Mahalanobis</i>	<i>N / T</i>	<i>BIAS</i>			<i>MSE</i>		
			<i>b</i>	$\tilde{\beta}(P = 1)$	$\tilde{\tilde{\beta}}(P = 1)$	<i>b</i>	$\tilde{\beta}(P = 1)$	$\tilde{\tilde{\beta}}(P = 1)$
$(\theta = 0, \tau = 42)$	0.0000	392 / 69	-0.0332	-0.0770	NA ^a	0.6667	1.2513	NA ^a
$(\theta = 0.5, \tau = 42)$	0.0400	392 / 63	0.0515	1.8935	1.3479	0.7690	4.8380	3.1545
$(\theta = 1.0, \tau = 45)$	0.1158	398 / 65	-0.0090	4.6333	2.2747	0.7911	22.9057	7.0200
$(\theta = 1.5, \tau = 45)$	0.1990	397 / 70	-0.0207	8.0477	3.1314	0.9097	66.2364	12.1340
$(\theta = 2.0, \tau = 48)$	0.3669	393 / 74	-0.0296	12.5921	11.5223	0.9633	160.1226	135.7373
$(\theta = 2.5, \tau = 65)$	0.3188	392 / 70	-0.0575	18.9455	11.9379	1.1112	360.4922	181.6960
$(\theta = 3.0, \tau = 75)$	0.4782	392 / 74	0.0142	22.8720	20.1200	1.2881	525.0096	407.0183

NOTE: This set of experiments corresponds to the “Set 2” experiments described in Table 2. “ θ ” is a parameter that adjusts the closeness of treatment and control observations. The experiment $\theta = 0$ represents the case of perfect matching. “ τ ” is a parameter that regulates the proportion of observations that are treatment observations. “Mean Mahalanobis” measures the average Mahalanobis distance between each treatment place and its matched control place (cf. equation (7) in the text). “N/T” reports the number of observations used in each replication of the CR and QECG estimations, respectively. “ b ” is the CR estimator; and $\tilde{\beta}(P = 1)$ and $\tilde{\tilde{\beta}}(P = 1)$ are the QECG estimators defined in equations (17) and (26) when each treatment place is matched to only one control place (i.e., $P = 1$). Table values are the mean values associated with 1000 replications of the respective experiment.

^a The QECG estimator $\tilde{\tilde{\beta}}$ is not defined in the case of perfect matching.

TABLE 5A
The Effect On CR And QECG Estimators Of Increasing Dissimilarity Between Treatment And Control Observations
When There Is Misspecification In The CR Model And $f(X)$ Is Nonlinear (“step”=2)

<i>Experiment</i>	<i>Mean Mahalanobis</i>	<i>N / T</i>	<i>BIAS</i>			<i>MSE</i>		
			<i>b</i>	$\tilde{\beta}(P = 1)$	$\tilde{\tilde{\beta}}(P = 1)$	<i>b</i>	$\tilde{\beta}(P = 1)$	$\tilde{\tilde{\beta}}(P = 1)$
$(\theta = 0, \tau = 42)$	0.0000	392 / 69	5.5083	-0.0770	NA ^a	30.9679	1.2513	NA ^a
$(\theta = 0.5, \tau = 42)$	0.0400	392 / 63	7.2887	1.8935	1.3479	53.8389	4.8380	3.1545
$(\theta = 1.0, \tau = 45)$	0.1158	398 / 65	13.3081	4.6333	2.2747	177.7892	22.9057	7.0200
$(\theta = 1.5, \tau = 45)$	0.1990	397 / 70	24.7594	8.0477	3.1314	613.6515	66.2364	12.1340
$(\theta = 2.0, \tau = 48)$	0.3669	393 / 74	32.6196	12.5921	11.5223	1064.6487	160.1226	135.7373
$(\theta = 2.5, \tau = 65)$	0.3188	392 / 70	49.7158	18.9455	11.9379	2472.3061	360.4922	181.6960
$(\theta = 3.0, \tau = 75)$	0.4782	392 / 74	65.9566	22.8720	20.1200	4350.8403	525.0096	407.0183

NOTE: This set of experiments corresponds to the “Set 3” experiments described in Table 2. The data underlying this table are identical to those in Table 4 except now the CR estimates are obtained from a misspecified regression. In particular, the estimated CR model does not allow for nonlinearity. The QECG results are identical to those in Table 4. Table values are the mean values associated with 1000 replications of the respective experiment.

^a The QECG estimator $\tilde{\tilde{\beta}}$ is not defined in the case of perfect matching.

TABLE 5B
The Effect On CR And QECG Estimators Of Increasing Dissimilarity Between Treatment And Control Observations
When There Is Misspecification In The CR Model And $f(X)$ Is Nonlinear ($step=1$)

<i>Experiment</i>	<i>Corr(D, f(X))</i>		<i>BIAS</i>			<i>MSE</i>		
	<i>step=2</i>	<i>step=1</i>	<i>b</i>	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$	<i>b</i>	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$
$(\theta = 0, \tau = 42)$	0.1803	0.0632	1.0921	-0.0180	NA	1.6851	0.8325	NA
$(\theta = 0.5, \tau = 42)$	0.1960	0.1365	4.0447	1.0190	0.5777	16.8628	1.9394	1.3211
$(\theta = 1.0, \tau = 45)$	0.3300	0.1731	5.8922	2.2933	0.9098	35.1864	5.9859	1.6999
$(\theta = 1.5, \tau = 45)$	0.5045	0.2437	9.6993	3.2464	0.8303	94.5394	11.3148	1.8147
$(\theta = 2.0, \tau = 48)$	0.5963	0.3233	16.5416	3.2985	1.4842	274.0453	11.6632	3.1331
$(\theta = 2.5, \tau = 65)$	0.6539	0.2880	19.2545	5.4772	1.7668	371.2210	30.9933	7.7111
$(\theta = 3.0, \tau = 75)$	0.7079	0.3337	27.5624	8.9213	6.2300	760.1937	80.5151	45.9174

NOTE: This set of experiments also corresponds to the “Set 3” experiments described in Table 2. The data generating process underlying this table is identical to that of Table 5A except that the “*step*” parameter is set equal to 1, rather than 2. This causes the treatment variable to be characterized by greater noise, reducing the correlation between D and $f(X)$. Table values are the mean values associated with 1000 replications of the respective experiment.

^a The QECG estimator $\tilde{\tilde{\beta}}$ is not defined in the case of perfect matching.

TABLE 6
The Effect On The QECG Estimator $\tilde{\beta}$ Of Increasing the Number of Control Places Matched To Each Treatment Place

<i>Experiment</i>	<i>Mean Mahalanobis</i>			<i>BIAS</i>			<i>MSE</i>		
	<i>P=1</i>	<i>P=3</i>	<i>P=5</i>	$\tilde{\beta}(P=1)$	$\tilde{\beta}(P=3)$	$\tilde{\beta}(P=5)$	$\tilde{\beta}(P=1)$	$\tilde{\beta}(P=3)$	$\tilde{\beta}(P=5)$
$(\theta = 0, \tau = 42)$	0.0000	0.0107	0.0147	NA ^a	0.2694	0.3530	NA ^a	0.9541	0.8704
$(\theta = 0.5, \tau = 42)$	0.0400	0.0671	0.0872	1.3479	1.3318	1.2857	3.1545	2.9365	2.8116
$(\theta = 1.0, \tau = 45)$	0.1158	0.1681	0.2076	2.2747	2.6869	3.1618	7.0200	9.1348	13.0045
$(\theta = 1.5, \tau = 45)$	0.1990	0.3247	0.4070	3.1314	4.8540	5.9107	12.1340	29.7091	45.7180
$(\theta = 2.0, \tau = 48)$	0.3669	0.4650	0.5433	11.5223	13.0746	14.0853	135.7373	172.0017	199.2774
$(\theta = 2.5, \tau = 65)$	0.3188	0.4868	0.5963	11.9379	21.1431	23.0230	181.6960	448.0685	530.9646
$(\theta = 3.0, \tau = 75)$	0.4782	0.6221	0.7672	20.1200	24.3965	28.2502	407.0183	596.5804	799.1671

NOTE: This set of experiments corresponds to the “Set 4” experiments described in Table 2. The data underlying this table are identical to those in Tables 4 and 5A except now the QECG estimator is calculated for the cases when treatment places are matched to multiple control places. Note specifically that the “Mean Mahalanobis”, “Bias” and “MSE” values when $P=1$ are identical to those in Tables 4 and 5A. Table values are the mean values associated with 1000 replications of the respective experiment.

^a The QECG estimator $\tilde{\beta}$ is not defined in the case of perfect matching.

TABLE 7
The Effect On CR And QECG Estimators Of Including Irrelevant Explanatory Variables
When $f(X)$ Is Nonlinear And There Is Misspecification In The CR Model

<i>Experiment: Variables included in CR Estimator and Used for Matching in QECG Estimators</i>	<i>Mean Mahalanobis (X_1 and X_2)</i>	<i>BIAS</i>			<i>MSE</i>		
		b	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$	b	$\tilde{\beta}(P=1)$	$\tilde{\tilde{\beta}}(P=1)$
X_1, X_2	0.0642	10.1185	3.6652	1.7435	103.1207	14.7256	4.7817
X_1, X_2, X_3	0.1346	10.0812	5.5238	2.8723	102.2903	31.6964	10.5286
X_1, X_2, \dots, X_4	0.2516	10.0277	6.2056	4.0272	101.2638	39.6606	18.0145
X_1, X_2, \dots, X_5	0.2969	10.1381	7.1069	5.4096	103.4814	51.6413	30.9192
X_1, X_2, \dots, X_6	0.2881	10.0992	6.6819	5.9857	102.6789	45.7960	37.1506
X_1, X_2, \dots, X_7	0.3025	9.8815	6.2899	5.7430	98.3863	40.8243	34.3757
X_1, X_2, \dots, X_8	0.3553	9.9975	6.7567	6.0840	100.6900	46.9184	38.4787
X_1, X_2, \dots, X_9	0.4266	9.9209	7.1985	6.2668	99.1363	53.0562	40.6021
X_1, X_2, \dots, X_{10}	0.4466	10.1294	7.4853	6.4638	103.2915	57.2011	43.1093
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_1, X_2, \dots, X_{50}	21.6654	9.7433	8.0212	6.8514	95.6376	65.4179	48.1579

NOTE: This set of experiments corresponds to “Set 5” described in Table 2. The true X consists of X_1 and X_2 . The parameters θ and τ are fixed at $(\theta = 1.0, \tau = 45)$ for all the experiments in the set. The individual experiments differ only in that they include additional (irrelevant) X ’s in X in both the CR and QECG estimation, where the irrelevant X ’s are generated from the same individual distributions as X_1 and X_2 , and are uncorrelated both with the relevant X ’s and each other. The “Mean Mahalanobis” values reported in the table are based on Mahalanobis distances using only characteristics X_1 and X_2 . These are distinct from the Mahalanobis values used in the QECG matching algorithm, which are based on all the characteristics in X . However, the latter are of little value in comparing across experiments since they increase monotonically with the number of included characteristics. Table values are the mean values associated with 1000 replications of the respective experiment.